

FINAL
7N-43-CR
9CIT.

FINAL STATUS REPORT
NASA GRANT NAGW-3924

**Research on Methods for Analysis of
Multispectral Earth Observational Data**

Principal Investigator

David A. Landgrebe
School of Electrical & Computer Engineering
Purdue University
West Lafayette, Indiana 47907-1285
Phone 317-494-3486
FAX 317-494-3358
Internet: landgreb@ecn.purdue.edu

Cognizant NASA Technical Officer

Dr. D. E. Wickland
Code YSE, NASA Headquarters

Year 3 of 3

For the research period February 1, 1996 to January 31, 1997

Summary

Objective. The work pursued under this grant has as its goal the definition of means for effectively analyzing multispectral and hyperspectral image data based upon principles of signal theory and modern signal processing technology in addition to physical and biophysical ones and the rapid availability of this technology to the Earth science research and practitioner community. This work is a logical extension of work done under NASA Grant NAGW-925.

Finding. The history of Landsat data analysis¹ shows that the data analysis problem is clearly a cross-disciplinary one. In more recent years, sensor technology has continued to advance, making more powerful and therefore more complex data available. This more complex data requires more sophisticated analysis methods to realize the full potential of such data. Thus a significant problem has developed in how to bridge the gap between the techniques required for such analysis and the background that most Earth scientists, the users of this technology, have in signal theory and signal processing fundamentals. Unless this cross-disciplinary gap can be bridged, the field must settle for analysis technology based upon simple cause/effect relationships between spectral characteristics and the desired information, methods which, while rather intuitive, are not state-of-the-art from a signal processing standpoint. Thus the need is to generate the required fundamental background signal processing knowledge and the algorithms for hyperspectral data analysis, and place them in a procedural context which will be found attractive to Earth scientists who are especially knowledgeable of the scene, but who are not specialists in signal processing technology.

Results. In our studies of methods for the analysis of multispectral and hyperspectral data over the last several years, the following not widely perceived fundamental concepts about the analysis of such data have become increasingly both apparent and pivotal.

- *As the number of features to be used in discriminating between ground cover types grows beyond three, the second order variations in the spectral response of a cover class become increasingly important compared to first order effects.* That is, the shapes of the class distributions in feature space become increasingly significant relative to discrimination between classes compared to their location in feature space. There are what appear at first to be counter-intuitive aspects to this. For example, two distributions may have the same mean value, and yet be separable. This is an extension of the fact that a single point in feature space, i.e., a single spectral response curve, does not adequately represent in a quantitative form the information available of a ground cover class.

¹ Landgrebe, David, "The Evolution of Landsat Data Analysis," (Invited), Photogrammetric Engineering and Remote Sensing, Vol. LXIII, No. 7, July 1997, to appear.

- *Related to this, as the number of features to be used in any given analysis is increased, the importance of increasingly precise quantitative characterization of the classes of ground cover of interest grows. This property has been known to signal processing specialists for some time, but has not been commonly recognized by Earth scientists. It points to the importance of accumulating as many design samples per class as possible, while at the same time reducing the number of features to be used to the smallest set which can provide the desired accuracy.*
- *The immediate goal for the discriminator design process for any given case is to establish an adequately precise, quantitatively defined list of classes which are simultaneously exhaustive, separable, and of informational value to the analyst. The implication here is that, though one is frequently interested in only one or a small number of classes in the scene, for optimal performance, all ground scene classes, i.e., the distribution of the entire data set in feature space, must be adequately modeled quantitatively.*

Previous work has focused upon devising means for optimal feature extraction to provide the smallest, yet most effective feature sets. Current results have been directed to testing the previously devised algorithms [1, 4]², devising straightforward analysis procedures for using these algorithms so that they will be more acceptable to the Earth science community [3, 8] and determining the optimal degree of classifier complexity for the design set available [5]. A method for decision tree classifier design was also examined as a means for focusing the feature design process on pairs of classes rather than attempting optimality over the entire list of classes [2]. A previously derived result on use of feature extraction in a neural network context also appeared during this period [9]. New results on the unique characteristics of high dimensional feature spaces and how they might be dealt with in the analysis process also were documented [10]. And finally, two papers describing aspects of the program to the Earth science user community were given [6, 7]

MultiSpec© Distribution

As stated above, in addition to the research itself, the intent is to deliver the knowledge gained and algorithms created in this research to the Earth science (user) community by as rapid and effective means as possible. Thus an integral part of the work has been to provide to this user community a capability to learn and evaluate the algorithms defined via an easy-to-acquire and easy to use software package. This software package is called MultiSpec©, and it has been under development in this mode and for this purpose for the last several years. The package has been copyrighted to protect against misuse, but it is freely distributed to requesters by the least expensive and most convenient means available. For the last two years, MultiSpec and its documentation has been downloadable from the world wide web at the following URL:

<http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/>

² Numbers in [] refer to papers and presentations listed below.

User interest has been significant. In previous years, several hundred requests for copies were received, and copies were mailed to these requesters. A summary of registered users up to May 31, 1994 is available at:

http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/Final_Report_table1.html

Although a more current summary would be much larger, this list reasonably illustrates the types of organizational units which are downloading copies during the current reporting period. Currently, copies are downloaded at a rate in excess of 200/month. New versions are posted on the net every few months, as new capabilities are added.

An unexpected past outcome of this means for technology transfer was that a substantial number of requests for MultiSpec came from teachers of the K-12 level. As a result of this type of interest, a distribution license has been granted to The Consortium for Mathematics and Its Applications (COMAP), a National Science Foundation funded program based in Lexington, Massachusetts, for use in their ARISE (Applications/Reform in Secondary Education) program. This is a 5-year project to generate new mathematics curricula for grades 9-11. They report having incorporated MultiSpec into their new curricula as a motivating tool. Many requests for MultiSpec have also come directly from secondary school teachers.

In the last two years, MultiSpec has become of interest to and has been adopted by the NASA/NOAA/NSF GLOBE program, (<http://www.globe.gov>) a program motivated by Vice President Gore to have K-12 children involved in collection of environmental data. More than 3500 schools across the U.S. and 52 foreign countries have signed up to participate. As a result, in addition to the primary Macintosh version, a Windows 3.1/Windows 95/Windows NT compatible version of MultiSpec is also in preparation.

A survey of the list of downloads indicates that approximately 40% were associated with GLOBE, ARISE, and other K-12 educational institutions. The other 60% are primarily from the remote sensing research, educational, and application community in the U.S and around the world, although a small number went into hospitals and other health care organizations.

Vegetation and Soils Field Research Data Summary

On a related note, we receive requests periodically for data from our Vegetation and Soils Field Research Data Base developed in the mid '70's to early '90s. The requesters are looking for spectral reflectance measurements of vegetation canopies and soil types under controlled conditions that are well documented. We have made a significant portion of this data available now to anyone having access to the WWW. The URL is:

<http://dynamo.ecn.purdue.edu/~frdata/FRData/>

The web pages contains summaries of the experiments for which spectral and agronomic data were collected. The data files for many of the experiments can be downloaded along with the documentation of the data formats.

Publications - Feb. 1, 1996 to Jan. 31, 1997

- [1] Xiuping Jia and David Landgrebe, "Large Area Classification For Hyperspectral Data Sets," 8th Australasian Remote Sensing Conference, Canberra, Australia, 25-29 March, 1996.
- [2] Saldju Tadjudin and David A. Landgrebe, "A Decision Tree Classifier Design For High-Dimensional Data With Limited Training Samples," Proceedings of the IGARSS '96 Symposium, Lincoln, NE, pp. 790-793, 27-31 May, 1996.
- [3] Pifuei Hsieh and David A. Landgrebe, "Automated Training Sample Labeling Using Laboratory Spectra," Proceedings of the IGARSS '96 Symposium, Lincoln, NE, pp. 1855-1859, 27-31 May, 1996.
- [4] J.A. Benediktsson, Kolbeinn Arnason, Arni Hjartarson, and David Landgrebe, "Classification And Feature Extraction Based On Enhanced Statistics," Proceedings of the IGARSS '96 Symposium, Lincoln, NE, pp. 414-416, 27-31 May, 1996.
- [5]* Hoffbeck, Joseph P. and David A. Landgrebe, "Covariance Matrix Estimation and Classification with Limited Training Data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, no. 7, pp. 763-767, July 1996.
- [6] David A. Landgrebe, "On the Information Content of Optical Land Remotely Sensed Data," (Invited), Progress In Electromagnetics Research Symposium, Innsbruck Austria, July 8-12, 1996.
- [7] Biehl, L. and David Landgrebe, "MultiSpec - A Tool for Multispectral-Hyperspectral Image Data Analysis," 13th Pecora Symposium, Sioux Falls, SD, August 20-22, 1996.
- [8]* Hoffbeck, Joseph P. and David A. Landgrebe, "Classification of Remote Sensing Images having High Spectral Resolution," *Remote Sensing of Environment*, Vol. 57, No. 3, pp. 119-126, September 1996.
- [9] Chulhee Lee and David A. Landgrebe, "Decision Boundary Feature Extraction for Neural Networks," *IEEE Transactions on Neural Networks*, Vol. 8, No. 1, pp. 75-83, January 1997.

Pending Publication

- [10]* Jimenez, Luis, and David Landgrebe, "Supervised Classification in High Dimensional Space: Geometrical, Statistical, and Asymptotical Properties of Multivariate Data," *IEEE Transactions on System, Man, and Cybernetics*, submitted December 5, 1995, Accepted August 29, 1996, To appear January, 1998.

* Papers marked * as well as several earlier papers from this work are available for downloading from the www at: <http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/documentation.html>